

Universidad del Sureste

BIOESTADISTICA

15/10/2021

Ensayo

Unidad 3



ALUMNA: Angela de Jesus Escobar Caballero

DOCENTE: Lic. Omar Ulises Hernández

4to Cuatrimestre



Ensayo

- **Muestreo aleatorio simple.**

Se puede decir que el muestreo aleatorio simple (M.A.S.) es la técnica de muestreo en la que todos los elementos que forman el universo y que por lo tanto están descritos en el marco muestral, tienen idéntica probabilidad de ser seleccionados para la muestra.

Un personaje como: César Pérez López, en su libro "Muestreo Estadístico" (Pearson, 2005) desarrolla de forma muy clara una comparación entre ambas técnicas. Tanto si lo miramos desde el punto de vista podemos saber de qué la técnica genera estimaciones más precisas como desde el punto de vista de igual manera qué la técnica permite tener la misma precisión con menor tamaño de muestra, se puede concluir que el muestreo aleatorio simple sin reposición siempre es más eficiente.

La fórmula relaciona el tamaño de muestra necesario cuando el universo es finito con el tamaño necesario cuando el universo es infinito.

Beneficios del muestreo aleatorio simple

El muestreo aleatorio simple es un procedimiento de muestreo probabilístico que da a cada elemento de la población objetivo y a cada posible muestra de un tamaño determinado, la misma probabilidad de ser seleccionado.

De igual manera el muestreo aleatorio simple no es tan utilizado en investigaciones del consumidor, sobre todo porque es complicado obtener un marco de muestreo donde extraer al azar y no querrás darle a todas las unidades de la muestra una probabilidad igual de ser elegidas, ya que usualmente para hacer una investigación de este tipo se requiere a usuarios de tiendas o consumidores de ciertos ~~productos o ciertas áreas específicas para ser las unidades de muestreo.~~

Pasos para seleccionar una muestra aleatoria simple

- Define la población objetivo. Quizá quieras leer: ¿Cómo encontrar a tu mercado objetivo?
- Identifica un marco de muestreo actual de la población objetivo o desarrolla uno nuevo.
- Evalúa el marco de muestreo para la falta de cobertura, cobertura excesiva, cobertura múltiple y la agrupación, y haz los ajustes que consideres necesarios.
- Asigna un número único a cada elemento de la trama.
- Determina el tamaño de la muestra.
- Selecciona al azar el número específico de elementos de la población.

También se puede decir ya que es de gran utilidad que el uso de números aleatorios, un método alternativo implica también la numeración de miembros de la población de 1 a N. Luego, el tamaño de muestra de n tiene que ser determinada por selección de los números al azar.

Este proceso de la tabla de números aleatorios es un proceso tedioso, consume tiempo, y no se recomienda para grandes poblaciones.

En su lugar, se pueden utilizar softwares estadísticos u hojas de cálculo para generar números aleatorios.

Subtipos de muestreo aleatorio simple

De igual manera existe dos tipos de muestreo aleatorio simple: el muestreo con reemplazo y sin reemplazo. En el muestreo con reemplazo, después de que un elemento ha sido seleccionado de entre el marco de la muestra se devuelve y es elegible para ser seleccionado de nuevo.

Ventajas y desventajas del muestreo aleatorio simple

- Entre sus puntos fuertes están que tiende a producir muestras representativas y permite el uso de la estadística inferencial en el análisis de datos recogidos.
- Cada selección es independiente de otras selecciones; Todas las combinaciones posibles de unidades de muestreo tienen la misma oportunidad de ser seleccionadas. En el muestreo sistemático, las posibilidades de ser seleccionado no son independientes entre sí.
- En general, es más fácil que otros procedimientos de muestreo probabilístico (tales como el muestreo por conglomerados) de comprender y comunicar a otros.
- Los procedimientos estadísticos requeridos para analizar los datos y calcular los errores son más fáciles que los requeridos en otros procedimientos de muestreo probabilístico. Te recomiendo leer: Tipos de errores. Consejos para que tu investigación no caiga en ellos

- Entre las desventajas están que se requiere un marco de muestreo de elementos de la población objetivo. Un marco de muestreo apropiado puede que no exista para la población que se dirige, y puede que no sea factible o práctico construir uno. En este caso el muestreo por conglomerados no requiere de una toma de muestra de los ~~elementos de la población objetivo.~~
- El muestreo aleatorio simple tiende a tener errores de muestreo más grandes y menos precisión de muestreo estratificado del mismo tamaño de la muestra.
- Los encuestados pueden estar muy dispersos, por tanto, los costos de la recolección de datos pueden ser superiores a las de otros diseños de la muestra de probabilidad, como el muestreo por conglomerados.
- El muestreo aleatorio simple puede no producir un número suficiente de elementos de pequeños subgrupos. Esto no haría de un muestreo aleatorio simple una buena opción para los estudios que requieren un análisis comparativo de las categorías pequeñas de una población con categorías mucho más amplias de la población.
- **Justificación del muestreo.**

En vez de tomar un censo completo, los procedimientos de muestreo estadístico se han convertido en la herramienta preferida en la mayoría de las situaciones de investigación.

Existen tres razones principales para extraer una muestra.

Antes que todo, por lo general, lleva demasiado tiempo realizar un censo completo. En segundo lugar, es demasiado costoso hacer un censo completo. Tercero, es demasiado molesto e ineficiente obtener un conteo completo de la población objeto

- **Función de Distribución empírica.**

Podemos decir que la distribución empírica es asociada a una muestra es la ley de probabilidad sobre el conjunto de las modalidades, que afecta a cada observación con el peso $1/n$.

La función de distribución empírica (FED) o cdf empírica es una función de paso que salta por $1/N$ a la ocurrencia de cada observación.

FED está estima la verdadera función de densidad acumulativa subyacente de los puntos en la muestra; Se garantiza virtualmente que converge con la distribución verdadera a medida que el tamaño de la muestra se hace lo suficientemente grande.

Este es la principal propiedad de la función de distribución empírica de la muestra es su aproximación a la función de distribución poblacional cuando aumenta el tamaño muestral.

- **Estadísticos muestrales. Distribuciones.**

En la estadística un estadístico (muestral) es una medida cuantitativa, derivada de un conjunto de datos de una muestra, con el objetivo de estimar o inferir características de una población o modelo estadístico.

Formalmente se sabe que un estadístico es una función medible T que, dada una muestra estadística de valores, les asigna un número, que sirve para estimar determinado parámetro de la distribución de la que procede la muestra.

Existen dos formas más comunes de estas variables corresponden a las distribuciones muestrales de las medias y de las proporciones.

Dada que en una población es constituida por un número n de elementos, cuya media aritmética es m y donde la desviación típica viene dada s , pueden formarse n^2 ~~muestras con reemplazamiento distintas, formadas por dos elementos de la~~ población.

De igual manera hay dos medidas fundamentales de esta distribución son la media y la desviación típica, también denominada error típico.

- **Estimación.**

La estimación puntual esta consiste en atribuir un valor (la estimación) al parámetro poblacional. Si esta muestra es representativa para la población, podemos esperar que los estadísticos calculados en las muestras tengan valores semejantes a los parámetros poblacionales, y la estimación consiste en asignar los valores de los estadísticos muestrales a los parámetros poblacionales.

Con los estadísticos con que obtenemos las estimaciones se denominan estimadores.

La estimación esta se puede o se divide en tres grandes bloques, cada uno de los cuales tiene distintos métodos que se usan en función de las características y propósitos del estudio:

Estimación puntual:

Método de los momentos;

Método de la máxima verosimilitud;

Método de los mínimos cuadrados;

Estimación por intervalos.

~~Estimación bayesiana.~~

Un estimador es como una regla que establece cómo poder calcular una estimación basada en las mediciones contenidas en una muestra estadística.

Estimación puntual

Esta consiste en la estimación del valor del parámetro mediante un sólo valor, obtenido de una fórmula determinada.

Se podría decir que estas son asignadas por letras y que sea insesgado (ausencia de sesgos) y estable en el muestreo o eficiente (varianza mínima).

Estimación puntual se representa ya Sea X una variable poblacional con distribución F_θ , siendo θ desconocido. El problema de estimación puntual consiste en, seleccionada una muestra X_1, \dots, X_n , encontrar el estadístico $T(X_1, \dots, X_n)$ que mejor estime el parámetro θ . Una vez observada o realizada la muestra, con valores x_1, \dots, x_n , se obtiene la estimación puntual de θ , $T(x_1, \dots, x_n) = \hat{\theta}$.

Estimación por intervalos

Este método consiste en la obtención de un intervalo dentro del cual estará el valor del parámetro estimado con una cierta probabilidad. En la estimación por intervalos se usan los siguientes conceptos:

Como intervalo de confianza

El intervalo de confianza es una expresión se identifica o se simboliza con del tipo $[\theta_1, \theta_2]$ ó $\theta_1 \leq \theta \leq \theta_2$, donde θ es el parámetro a estimar. Este intervalo contiene al parámetro estimado con un determinado nivel de confianza. Pero a veces puede ~~cambiar este intervalo cuando la muestra no garantiza un axioma o un equivalente~~ circunstancial.

- **Propiedades de los estimadores.**

ESTIMADOR:

Este se le conoce como un estadístico (es decir, es una función de la muestra) usado para estimar un parámetro desconocido de la población. Por ejemplo, si se desea conocer el precio medio de un artículo (el parámetro desconocido) se recogerán observaciones del precio de dicho artículo en diversos establecimientos (la muestra) y la media aritmética de las observaciones puede utilizarse como estimador del precio medio.

SESGO:

Este se le denomina sesgo de un estimador a la diferencia entre la esperanza (o valor esperado) del estimador y el verdadero valor del parámetro a estimar.

EFICIENCIA:

Un estimador es más eficiente o preciso que otro, si la varianza del primero es menor que la del segundo.

CONVERGENCIA:

Para poder estudiar las características de un estimador no solo basta con saber el sesgo y la varianza, sino que además es útil hacer un análisis de su comportamiento y estabilidad en el largo plazo, esto es, su comportamiento asintótico. Podemos construir sucesiones de estimadores y estudiar el fenómeno de la convergencia.

- **Obtención de estimadores.**

Se puede observar que esta trata de un método de obtención de estimadores muy intuitivo. Básicamente, consiste en igualar los momentos poblacionales (que sean función del o los parámetros a estimar) con los momentos muestrales y despejar el parámetro al estimar.

Ventajas y desventajas del método de los momentos

Este método de los momentos es bastante sencillo y brinda estimadores compatibles (debajo suposiciones muy débiles), aunque estos estimadores son a menudo sesgados.

De igual manera se sabe que en algunos casos, cuándo estimamos parámetros de una familia conocida de distribuciones de probabilidad, este método es sustituido el método de máxima verosimilitud de Fisher, porque con máxima verosimilitud los estimadores tienen probabilidad más alta de ser cercanos a las cantidades que estimamos y son menos sesgadas.

- **Estimación por intervalos de confianza.**

Este es un método llamado estadística inferencial la cual tiene como fin el proceso de uso de los resultados derivados de las muestras para obtener conclusiones acerca de

las características de una población. La estadística inferencial esta nos permite estimar características desconocidas como la media de la población o la proporción de la población.

Una estimación del intervalo de confianza es un rango de números, llamado intervalo, ~~construido alrededor de la estimación puntual. El intervalo de confianza se construye~~ de manera que la probabilidad del parámetro de la población se localice en algún lugar dentro del intervalo conocido.

La idea de Intervalo de Confianza, es proponer un rango de valores entre los que posiblemente se encuentre el verdadero valor del parámetro θ .

- **Contraste de hipótesis.**

Dentro de la inferencia estadística, esta es un contraste de hipótesis (también denominado test de hipótesis o prueba de significación) es un procedimiento para juzgar si una propiedad que se supone en una población estadística es compatible con lo observado en una muestra de dicha población.

Esta terminación es fuertemente asociada al concepto estadístico de potencia y a los conceptos de errores de tipo I y II, que definen respectivamente, la posibilidad de tomar un suceso falso como verdadero, o uno verdadero como falso.

Las hipótesis pueden clasificarse en dos grupos, según:

- Especifiquen un valor concreto o un intervalo para los parámetros del modelo.
- Determinen el tipo de distribución de probabilidad que ha generado los datos.

Como por ejemplo del primer grupo es la hipótesis de que la media de una variable es 10, y del segundo que la distribución de probabilidad es la distribución normal.

POSIBLES ERRORES EN EL CONTRASTE DE HIPÓTESIS

El contraste de hipótesis no establece la verdad de la hipótesis, sino que un criterio que nos permite decidir si una hipótesis se acepta o se rechaza, o el determinar si las muestras observadas difieren significativamente de los resultados esperados.

En este proceso podemos incurrir en dos tipos de errores según sea la situación real y la decisión que tomemos.

- **Concepto y definiciones.**

Una hipótesis estadística: es una afirmación respecto a alguna característica de una población. La cual hace contrastar una hipótesis es comparar las predicciones con la realidad que observamos.

La hipótesis emitida: Esta se suele designar por H_0 y se llama Hipótesis nula porque parte del supuesto que la diferencias entre el valor verdadero del parámetro y su valor hipotético es debida al azar, es decir no hay diferencia.

La hipótesis contraria: se designa por H_1 y se llama Hipótesis alternativa.

Los contrastes: estos pueden ser unilaterales o bilaterales (también llamados de una o dos colas) según establezcamos las hipótesis, si las definimos en términos de igual y distinto estamos ante una hipótesis unilateral, si suponemos una dirección (en términos de mayor o menor) estamos ante uno unilateral.

- **Construcción de Test de hipótesis.**

Seis pasos básicos para configurar y realizar correctamente una prueba de hipótesis.

1. Especificar las hipótesis.
2. Elegir un nivel de significancia (también denominado alfa o α).
3. El gerente selecciona un nivel de significancia de 0.05, que es el nivel de significancia más utilizado.
4. Determinar la potencia y el tamaño de la muestra para la prueba.
5. Recolectar los datos.
6. Comparar el valor p de la prueba con el nivel de significancia.
7. Después de realizar la prueba de hipótesis, el gerente obtiene un valor p de 0.004. El valor p es menor que el nivel de significancia de 0.05.
8. Decidir si rechazar o no rechazar la hipótesis nula.

- **Contraste de hipótesis paramétricas.**

El problema del contraste de hipótesis es que este consiste básicamente en comprobar cotejar, decidir, en definitiva, sobre la veracidad de una hipótesis prefijada previamente como supuestamente cierta.

- **Test para poblaciones normales.**

Es una prueba de 2 muestras se puede utilizar para comparar si las medias de dos grupos independientes son diferentes. Esta prueba se deriva bajo el supuesto de que ambas poblaciones están normalmente distribuidas y poseen varianzas iguales. Si bien el supuesto de normalidad no es crítico (Pearson, 1931; Barlett, 1935; Geary, 1947), el supuesto de varianzas iguales es crítico si los tamaños de las muestras son notablemente diferentes (Welch, 1937; Horsnell, 1953).

- **Test para poblaciones binomiales y de Poisson.**

Nos encontramos con un modelo derivado de un proceso experimental puro, en el que se plantean las siguientes circunstancias.

Se realiza un número n de pruebas (separadas o separables).

Cada prueba puede dar dos únicos resultados A y \bar{A}

La probabilidad de obtener un resultado A es p y la de obtener un resultado \bar{A} es q , con $q= 1-p$, en todas las pruebas.

Esto implica que las pruebas se realizan exactamente en las mismas condiciones y son, por tanto, independientes en sus resultados. Si se trata de extracciones, (muestreo), las extracciones deberán ser con devolución (reemplazamiento) , o bien población grande (M.A.S).

- **Test basado en el estadístico chi-cuadrado.**

Una prueba de chi-cuadrada es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos.

Existen varios tipos de pruebas de chi-cuadrada:

~~Prueba de bondad de ajuste de chi-cuadrada~~

Como por ejemplo utilice este análisis para probar qué tan bien una muestra de datos categóricos se ajusta a una distribución teórica.

De igual manera usted puede realizar esta actividad y comprobar si un dado es justo, así mismo lanzando el dado muchas veces y utilizando una prueba de bondad de ajuste de chi-cuadrada para determinar si los resultados siguen una distribución uniforme.

- **Test de bondad de ajuste.**

La bondad de ajuste de un modelo estadístico describe lo bien que se ajusta un conjunto de observaciones. Las medidas de bondad en general resumen la discrepancia entre los valores observados y los valores esperados en el modelo de estudio. Tales medidas se pueden emplear en el contraste de hipótesis, e.g. el test de normalidad de los residuos, comprobar si dos muestras se obtienen a partir de dos distribuciones idénticas (ver test de Kolmogorov-Smirnov), o si las frecuencias siguen una distribución específica (ver ji cuadrada).

En el análisis de regresión, podemos ver que existen los siguientes indicadores:

- Coeficiente de determinación (El R^2 mide la bondad de ajuste)
- Lack-of-fit sum of squares.

En los sistemas reales regularmente nos encontramos con variables cuyo comportamiento es aleatorio, y son susceptibles de ser modeladas por variables de ~~entrada de un modelo estocástico, estas variables requieren de un tratamiento~~ estadístico para su generación de manera artificial, el cual se realiza usualmente por medio de un modelo teórico de distribución de probabilidad.

- **Test de heterogeneidad.**

Cuando hablamos de heterogeneidad podemos distinguir los dos aspectos: por un lado el relativo a las diferencias existentes entre los estudios en cuanto a características de los pacientes incluidos, la metodología utilizada, el tiempo de seguimiento, las dosis empleadas, la localización geográfica, etc. y por otro lado el concepto de heterogeneidad estadística ,que únicamente esta cuantifica la variabilidad entre los resultados de los estudios, y que puede ser debida a las diferencias reales de planteamiento y ejecución entre los estudios incluidos, o a otras causas.

- **Test de homogeneidad.**

Se plantea en este problema de la existencia de homogeneidad entre r poblaciones, para lo cual se realizan muestras independientes en cada una de ellas. Los datos mostrables vienen clasificados en s clases y sus frecuencias absolutas se presentan en forma de una matriz $r \times s$. siendo n_{ij} el número de observaciones en la i -ésima población pertenecientes a la j -ésima clase.

Este se quiere contrastar la hipótesis nula de que las probabilidades asociadas a las s clases son iguales en las r poblaciones. Donde n_i es el tamaño muestral para la i -ésima población, n_j es la frecuencia marginal de la j -ésima clase y n es el tamaño muestral total.

El estadístico L se distribuye como una χ^2 con $(r - 1)(s - 1)$ grados de libertad.

El contraste se realiza con un nivel de significación del 5%.

- **Tablas de Contingencia.**

En estadística estas son las tablas de contingencia se emplean para registrar y analizar la asociación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales).

Una tabla de contingencia es una tabla que cuenta las observaciones por múltiples variables categóricas. Las filas y columnas de las tablas corresponden a estas variables categóricas.

De manera formal, se simboliza o se consideran X e Y dos variables categóricas con I y J categorías respectivamente. Una observación puede venir clasificada en una de las posibles $I \times J$ categorías que existen.

De igual manera una tabla de contingencia (o tabla de clasificación cruzada), con I filas y J columnas se denomina una tabla $I \times J$.

La tabla de contingencia es un medio particular de representar simultáneamente dos caracteres observados en una misma población, si son discretos o continuos reagrupados en clases. Los dos caracteres son el tamaño de la muestra.

Bibliografía

<https://plataformaeducativauds.com.mx/assets/docs/files/asignatura/9dc9876863afc854013788efd7ec3825.pdf>